Bacteria Classification via Surface Enhanced Raman Spectroscopy and Principal Component Analysis

J. Guicheteau, S. Christesen, D. Emge, A. Hyre, L. Argue
U.S. Army Edgewood Chemical Biological Center
Aberdeen Proving Ground, MD, USA 21010-5424

Abstract

Surface-enhanced Raman scattering (SERS) provides rapid fingerprinting of biomaterial in a non-destructive manner. The adsorption of colloidal silver to biological material suppresses native biofluorescence while increasing the normal Raman signal via the surface-enhanced Raman effect. This work validates the applicability of qualitative SER spectroscopy, utilizing principal component analysis (PCA) to show discrimination of biological threat simulants, based upon multivariate statistical confidences limits bounding known data clusters. Several strains of *Bacillus* spores are investigated along with *Pantoea agglomerans*,and *Brucella noetomae*.

1. Introduction

The need for faster, more reliable biological detection systems is greater now then ever before as world wide affairs have raised an alarm in our collective consciousness concerning Chemical, Biological, Radiological, Nuclear, Explosive (CBRNE) threats. In contrast to their nuclear counterparts, biological weapons do not require hard-to-acquire infrastructure or base materials, but their effects can be just as devastating with infective doses ranging from approximately 8,000 organisms for *Bacillus anthracis* to only a few organisms for *Poxvirus variolae*, the causative agent for smallpox (Henderson; Inglesby et al. 1999: Peters; Hartley 2002). In the US alone, occurrences such as the 2001 anthrax mailings and the 1984 salmonella poisonings of salad bars in Dalles, Oregon have focused the populations fears on the possible large-scale release of deadly bacteria or viruses(Torok; Tauxe et al. 1997). Clearly, the need for biological detection capability has spread beyond exclusive military operations and soldier protection and into the arena of civilian population protection as well. More and more first responders such as fireman and police personnel will be called on to enter possible contaminated areas and perform rapid and accurate analyses to ascertain threat levels. Therefore, the need for real-time detection and identification of biological threat materials is imperative, and surface-enhanced Raman spectroscopy (SERS) has the potential to provide such a capability.

Following the discovery of surface-enhanced Raman (SER) in the mid 1970's in which a phenomenally strong enhancement of the normal Raman signal was observed from pyridine molecules adsorbed on roughened silver electrodes (Albrecht; Creighton 1977: Fleischmann; Hendra et al. 1974), the technique has been increasingly applied to more complex molecules and biologically relevant materials (Cotton; Schultz et al. 1980). As SERS becomes more efficient and reliable as a whole-organism fingerprinting technique greater effort to develop SERS based biological detection schemes are being investigated.

Recently biological SERS studies have incorporated the use of multivariate analysis for positive identification and discrimination between species. Jarvis and coworkers demonstrated the discrimination of bacterial isolates associated with urinary tract infection and strain level discrimination of *Escherichia coli* (Jarvis; Brooker et al. 2004). Additional Raman and SERS studies have demonstrated discrimination of pathogens (Gonser; Christesen et al. 2004: Huang; Griffiths et al. 2004), and showed classification/discrimination of both chemical and biological simulants (Pearman; Fountain 2006).

Through these earlier works, SERS has been established as a rapid and highly specific identification and detection technique for bacteria. This is partly due to SERS being amenable to the aqueous environments typical of biological systems as a result of the relatively small Raman scattering cross section of water (McCreery 2000). Furthermore, metal substrates used for SERS helps suppress the fluorescence background that more often than not plagues biological Raman experiments while at the same time providing an enhancement of the normal Raman signal by a factor 10^4 or greater.

Typical SERS-active substrates are metallic structures consisting of roughened noble metal surfaces such as gold and silver that exhibit unique free electron optical properties due to their $d^{10}s^1$ electronic configurations (Mullen; Wang et al. 1992). In this work the use of silver nanoparticles produced via a modified version of the Lee and Meisel preparation as described below were employed.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Bacteria Classification via Surface Enhanced Raman Spectroscopy and Principal Component Analysis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Edgewood Chemical Biological Center Aberdeen Proving Ground, MD, USA 21010-5424**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The primary focus of this paper is to further explore discrimination of gram positive and gram negative bacteria representing possible biological threat agents of interest to the US military utilizing principal component analysis (PCA) and SERS. *Bacillus anthracis sterne* and *Bacillus anthracis ames* (formalin killed) are non-pathogenic and pathogenic strains of *Bacillus anthracis*, respectively. *Bacillus thuringensis, and Bacillus globigii* are non-pathogenic surrogates for *Bacillus anthracis*. *Pantoea agglomerans* is a non-spore forming bacterial simulant for *Pasturella tularensis*, the causative agent for tularaemia and *Brucella neotomae* is a subtype species of *Brucella melitensis* which is a food pathogen that can affect cattle and bison but is non-pathogenic to humans.

2. Experimental Section

2.1 **Bacterial Strains**

Bacterial simulants were provided by the Armed Forces Institute of Pathology (AFIP) and the Biodefense (BD) Team of the US Army Edgewood Chemical Biological Center. Spores of *Bacillus anthracis ames* and *Bacillus anthracis sterne* (live and killed) were received in 15 mL of deionized H_2O from AFIP at concentrations of $1.19x10^9$ and $1.75x10^{10}$ respectively. These suspensions were stored at 4°C until needed. Washed vegetative cells of *Pantoea agglomerans* (Formally known as *Erwinia herbicola*-ATCC 13329), *Brucella neotomae* (ATCC 23459) and spores of *Bacillus subtilis globigii* (Dugway Proving Grounds-old) were received from AFIP in lyophilized form and stored at 4°C until needed. Finally, a suspension of *Bacillus thuringensis* spores were supplied by the ECBC BD team and had a concentration of $8.33x10^8$ cfu/mL.

Lyophilized samples of the bacteria were reconstituted by injecting 2 milliliters of deionized water (18.3 MΩ) into each vial followed by vortexing for 10 seconds to allow complete re-suspension. Samples were diluted to final concentrations of $1.08x10^9$ spores/mL of *Bacillus globigii*, $1.0x10^{10}$ cfu/mL of *Pantoea agglomerans*, and $4.87x10^{11}$ cfu/mL of *Brucella neotomae*.

Aluminum-coated microscope slides (ThermoElectron) were labeled as to the specific bacterium being investigated and prepared as follows. For each bacterium analyzed, 3 μL of colloidal silver were deposited followed by injection of 2μl of the bacterium. This procedure was repeated for each bacterium to provide duplicate samples. Slides were then placed in a desiccator and allowed to dry.

2.2 **Silver Colloid Preparation**

Silver nanoparticle suspensions were prepared following a modified procedure of Lee and Meisel (Lee; Meisel 1982). In a one-liter three neck round bottom flask, 90 mg of silver nitrate (99+ %, Aldrich) were dissolved in 400 mL of deionized water (18.3 MΩ). In order to control volume loss due to evaporation a condenser (250 mm jacket length) was placed on the center neck of the flask. A 25 mL addition funnel, with 10 mL of a 1% aqueous sodium citrate solution was attached to a second neck of the flask. A heating mantle was controlled by a 140-volt power-stat to bring the silver nitrate solution to a controlled boil in approximately 24 minutes. A 1" stir bar was placed in the reaction flask and set to stir at a rapid pace. In addition to above-mentioned set up, a heating top was placed around the flask to maintain temperature stability. Upon boiling, sodium citrate was added rapidly (stop cock fully opened). Heating and stirring were maintained for 45 minutes after addition of citrate, at which time the reaction flask was removed from heat and allowed to cool to room temperature (with stirring continued). The resulting nanoparticle suspensions appeared yellowish-brown with an electronic absorption λ_{max} of 400 nm, an average full width half maximum of 65 nm, and an average particle size of 35 nm.

2.3 Raman Instrumentation

A Bruker Optics Senturion Raman Microscope operating at 785 nm was used. The system allowed the laser to be focused onto a sample and data collection occurred over a full spectral range (200-3200 cm^{-1}) with a resolution of 8-10 cm^{-1}. The microscope also has a video attachment, allowing visual observation of the samples at powers of magnification up to 60x. Therefore, a very specific area of the sample could be targeted with the 5 μm laser spot. Spectra were collected at 40x magnification with 20-second integration times. Laser power was adjusted below 5.3 mW to avoid damage to the samples.

2.4 Data Collection

Data were collected from each sample spot following an "S" type pattern. A starting area was selected and an initial spectrum acquired, after which, 120 spectra were collected throughout the sample as to get a good representation of the entire SER-bacterial spot.

2.5 Data Processing.

Processing of SER spectral data is both a complex science and a bit of applied intuition. The intuitive nature of the analysis being made is a necessity due to

the complex and sometimes random nature of SER spectra itself as well as the subtle nuances associate with collection of those spectra. This results in a requirement that the data be normalized to compensate for systematic errors and screened for anomalies. The overall processing of the spectra in this work is to remove low order, 'baseline', trends, and then screen the data to remove anomalous behavior. The resulting data set is then transformed to a more useful space via Principle Component Analysis (PCA) where it is clustered in Principle Component (PC) space to isolate and identify groupings by specific species.

Before analyzing any Raman based spectra the removal of baselines or low order trends needs to be performed. These low order or slowly varying baselines are the result of surface reflections, florescence etc.. The selected method of baseline removal is based upon an iterative median filter technique. The median filter applied to the data was designed to remove spectral features with a base width greater than 300 cm^{-1}. Where the base width is defined as the distance between the local minima points that define a smoothed peak. This cutoff width was selected based upon observations of biological Raman signatures for various species including (*B. globigii* and *P. agglomernas*).

The implications and effects of the median filter are discussed in detail in Gallagher and Wise (Gallagher; Wise 1981). Using multiple iterations of the median filter, results in increased robustness when dealing with the confounding effects of both short constant neighborhoods, regions of monotonicity, and interactions within the median window of impulse and oscillation trends. In their work, Gallagher and Wise describe the median filter in a theoretical framework and show that a signal can be reduced to a single so-called root function which is invariant under additional applications of the same median filter. By adjusting the width of the median filter, the pass band of the filter can be adjusted to remove only signals that are not of interest to the current methodology. The poor behavior, as noted above, of real world signals is the driving reason for the multiple applications of the median filter. The use of additional iterations has shown an increase in the robustness of the filter to exception of assumed constant neighborhoods discussed in Gallagher and Wise, as well as improved performance when dealing with impulse like and oscillatory trends.

Before any analysis can be initiated concerns over data integrity of the collected spectra must be addressed. Anomalous data have been noted in the collected SER spectra and appear as dramatically differing baselines and overly broad features. Both of these phenomena are readily apparent to the researcher as well as the data analyst upon inspection. The most common cause of these anomalous spectra was improper focus on the surface of the material due to variations in sample deposition thickness. A systematic method to screen for these anomalies was implemented based on the cross-correlation over a population of samples from a given species. A qualitative review of the spectra from various species as well as ensuing statistical analysis lead to the observation that "good" data have cross-correlations greater than 0.65. Therefore by calculating the cross-correlation within each species and applying a lower threshold of 0.50 one can identify radically anomalous spectra. The identified spectra were removed from the data set for a given species. In total 770 spectra were collected and only 20 spectra fell below the 0.50 threshold.

The remaining SER spectra contain a myriad of redundant information which is known to mask the unique content that defines various species. To remove this masking and access the information necessary to uniquely identify individual bacteria the application PCA has been utilized. As applied, PCA is a data reduction and analysis technique that allows spectral similarities or differences to be easily seen. If the spectra of the biological material have definite, repeatable similarities, the data points from each sample should cluster together, showing successful discrimination of a sample-specific SERS signature. In this work the accepted method of eigen-decomposition of the correlation matrix was implemented to extract the PCs. This decomposition leads to the transformation of the data into a PC space with axes in the directions of maximal and orthogonal variation in the data. It is assumed that the directions of maximum variance will yield the greatest seperability of different species or classes within the given data set.

With the data transformed into the new PC space the need to locate, isolate, describe and represent the clusters formed by the data is needed. Since the first 3 principle components accounted for greater than 90% of the variance in the analyzed SER spectra all ensuing analysis was performed in the space defined by those three components. After decomposition of the data into PCA space, the data were checked for statistical integrity across spectra associated within each species. Using the a-priori knowledge of species cluster membership, the centroid of each species cluster was located based upon Euclidean distance metrics. With the centroid identified the Mahalanobis distance of each member of the species cluster was then calculated. Based on the Adaptive Mixture Model developed by Priebe (Priebe 1994) and the extensions to trivariate case by Sloka (Solka; Poston et al. 1995) a cutoff threshold of 3.54 was applied for data representation. Based upon Martinez and Martinez (Martinez; Martinez 2001) this corresponds to roughly a 2 standard deviation cutoff, adjusted for dimensionality, thereby removing statistical outliers. With the removal of

the statistical outliers new species clusters were formed and presented.

3. Results & Discussion

The graphical representations of the reduced clusters in PCA1 through PCA3 space are presented with corresponding confidence limits for each species cluster, based upon normally distributed data. These limits, or variance boundaries, were calculated for each individual species cluster to show the statistical space representing that species in the reduced PCA space. Furthermore, the following PCA plots are shown with reduced data sets for graphical clarity, however, all data points are shown in the constructed dendrograms.

Due to the rich informational content found in the regions between 500 and 1800 cm-1, PCA analyses were performed exclusively in this area. PCAs incorporating the strong CH stretching vibrations (2950–3100 cm^{-1}) did not produce improved differentiation and thus were not used in this work

3.1 PCA of *Bacillus* spores

Representative SER spectra of *B. globigii* (Bg), *B. anthracis ames* (Baa), *B. anthracis sterne*, live and killed (Basl & Bask), and *B. thuringensis* (Bt) are shown in Figure 1. While they do all share similarities in their respective spectra, several key variations are observed in the strains which aid in their PCA discrimination and are better described after viewing the constructed PCA.

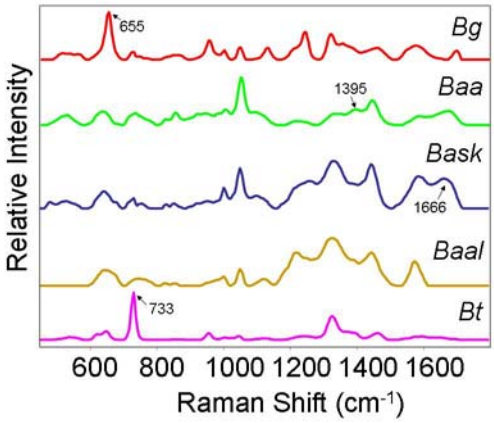

Figure 1. Representative SER spectra for five *Bacillus* strains.

Figure 2 shows a PCA plot of the first three PCs for the five *Bacillus* strains with statistical confidence limits shown for 1σ, 2σ, and 3σ. As can be seen, five distinct clusters are observed showing good discrimination between the bacteria in this space. These clusters are somewhat expected because the five spectra display observable differences. Both spores of *B. globigii* and *B. thuringensis* show significant spectral differences from the three anthracis strains and each other and are thus clustered in different areas of PC space. The primary spectral difference occur in the Bg band at 655 cm^{-1} and the Bt mode at 733 cm^{-1}. As for the anthracis strains being somewhat less separated, again referring to their spectra will explain this. The two *anthracis sterne* samples are the most similar however, the peak at 1666 cm^{-1} in the formalin killed *sterne* is noticeably absent in the live sterne spectrum. This small difference is enough to divide the two strains in PC space. *B. anthracis ames* again shares similarities with the *sterne* samples, and especially the killed bacteria, however the weak mode at 1395 cm^{-1} and other peak ratio variances is enough to distinguish the two strains.

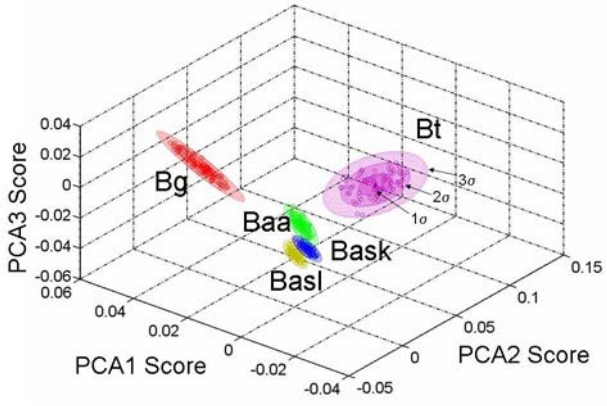

Figure 2. PCA of *Bacillus* strains using 1st three principal components.

To further show the discrimination between the five *Bacillus* species a dendrogram was constructed based on the mahalanobis distance between groups of scores for the first three PCs (Figure 3). From the *Bacillus* PCA, the three anthracis strains were more spectrally similar to each other than to either Bg or Bt. The dendrogram however, clearly shows good separation of the *Bacillus anthracis* strains as well.

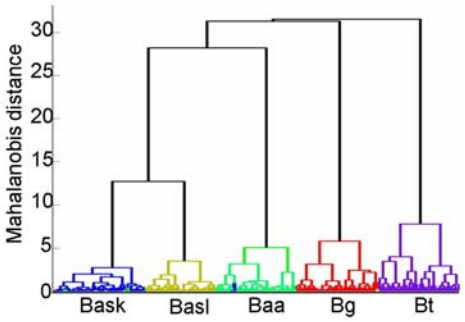

Figure 3. Dendrogram showing 5 distinct groups of *Bacillus* strains.

3.2. PCA of Various Bacteria

Following initial discrimination between the *Bacillus* strains, data for two gram negative bacteria were collected and incorporated into a PCA construct. Representative SER spectra of *P. agglomerans* (Pa) and *B. neotomae* (Bn) are shown in Figure 4. Again there are clear differences between the two bacteria and even more so when compared to the *Bacillus* samples.

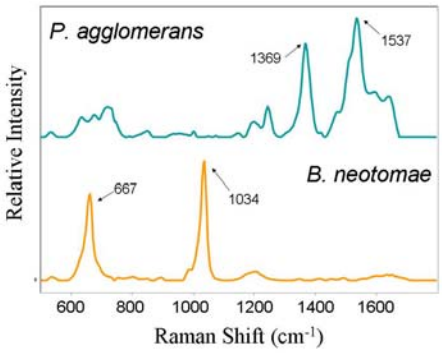

Figure 4. Representative SER spectra of two gram-negative bacteria.

In comparing *P. agglomerans* to the *Bacillus* strains the most significant spectral differences come from the 1369 cm^{-1} and 1537 cm^{-1} modes, respectively. For *B. neotomae*, the two main vibrational bands of 667 cm^{-1} and 1034 cm^{-1} distinguish it from the *Bacillus* samples. The PCA of the two gram negative bacteria and the five gram positive bacteria is shown in figure 5.

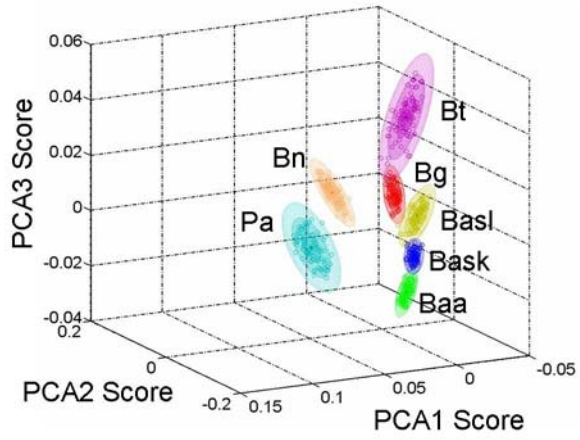

Figure 5. PCA of first three PC components showing good classification between gram-negative and gram-positive bacteria.

As can be seen *P. agglomerans* and *B. neotomae* readily separate from the five *Bacillus* species. Again, due to the already stated spectral differences this is understood. However, from the angle of the current axis of figure 5, it appears that in PC space the two gram-negative species are not far removed from the *Bacillus* species. Nevertheless, if the angle of the plot is rotated a very different depiction can be seen. Figure 6, shows the same plot as figure 5, but rotated to show only the first two principal components. Now what is easily observed is that in fact *B. neotomae* and *P. agglomerans* are far removed from one another and the now almost singularly clustered *Bacillus* strains.

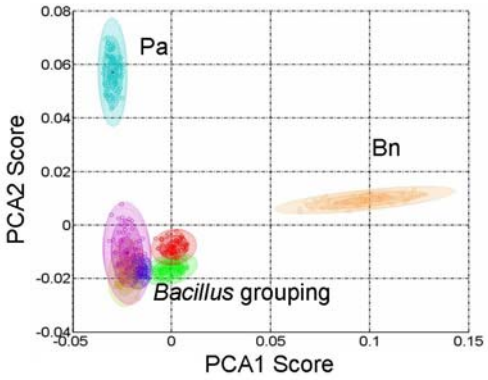

Figure 6. 2D plot showing greater separation between gram-negative and gram-positive bacteria.

While rotating the PC to show greater classification between the gram negative species, it consequently reduced, visually at least, the seperation between the *Bacillus* samples. However, when the dendrogram is constructed (Figure 7) clear classification between the gram negative species are seen as well as the probable discrimination within the *Bacillus* genome.

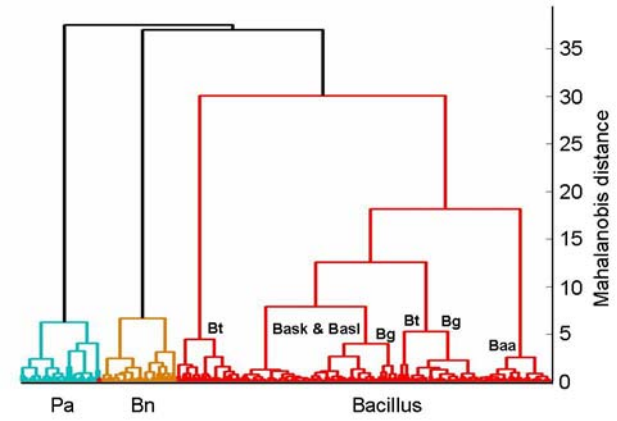

Figure 7. Dendrogram showing discrimination between gram-negative and gram positive bacteria.

We say "probable discrimination" due to the fact that even though five groups can be depicted in the *Bacillus* dendrogram cluster, there is quite a bit of spectral mixing of the strains. The fartherst left cluster of the *Bacillus* class is mostly made up of *B. thuringensis* spectra. However a small group of samples are intermixed between *B. anthracis sterne* live cluster and *B. gloibgii* groups. Although *B. anthracis ames* shows good speration from the the other strains, there was some mixing with other *B. anthracis* strains. We believe, that

with futher refinement of this technique, an improved classification method can be obtained.

4. Conclusion/Future Work

The results presented clearly show the ability to discriminate between *Bacillus* spores using SERS coupled with principal component analysis. We also show the differentiation of bacteria in larger data sets that include gram negative and gram positive bacteria. Further refinement of the multivariate analysis application for classification of bacteria is needed to produce a more authoritative technique. Problems that need to be addressed are differences in growth media, laboratory matriculation, and bacteria that have undergone washings to remove excess growth media. We are currently acquiring SER spectra of additional bacteria while investigating other multivariate analysis techniques such as modified PCA, discriminate function analysis and independent component analysis for improved identification and classification. The goal is to develop algorithms that provide bacterial identification and detection of unknowns using surface-enhanced Raman spectra.

References

Albrecht, M. G. and J. A. Creighton, 1977: Anomalously intense Raman spectra of pyridine at a silver electrode. *Journal of the American Chemical Society*, **99,** 5215-17.

Cotton, T. M., S. G. Schultz, and R. P. Van Duyne, 1980: Surface-enhanced resonance Raman scattering from cytochrome c and myoglobin adsorbed on a silver electrode. *Journal of the American Chemical Society*, **102,** 7960-2.

Fleischmann, M., P. J. Hendra, and A. J. McQuillan, 1974: Raman-Spectra of Pyridine Adsorbed At a Silver Electrode. *Chemical Physics Letters*, **26,** 163-166.

Gallagher, N., Jr. and G. Wise, 1981: A theoretical analysis of the properties of median filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **29,** 1136-1141.

Gonser, K. R., S. D. Christesen, and J. A. Guicheteau, 2004: Raman spectroscopy of biological material: Results from the RAAD common sample set phase II. *Proceedings of SPIE-The International Society for Optical Engineering*, **5585,** 136-142.

Henderson, D. A., T. V. Inglesby, J. G. Bartlett, M. S. Ascher, E. Eitzen, P. B. Jahrling, J. Hauer, M. Layton, J. McDade, M. T. Osterholm, T. O'Toole, G. Parker, T. Perl, P. K. Russell, and K. Tonat, 1999: Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA: the journal of the American Medical Association*, **281,** 2127-37.

Huang, W. E., R. I. Griffiths, I. P. Thompson, M. J. Bailey, and A. S. Whiteley, 2004: Raman Microscopic Analysis of Single Microbial Cells. *Analytical Chemistry*, **76,** 4452-4458.

Jarvis, R. M., A. Brooker, and R. Goodacre, 2004: Surface-Enhanced Raman Spectroscopy for Bacterial Discrimination Utilizing a Scanning Electron Microscope with a Raman Spectroscopy Interface. *Analytical Chemistry*, **76,** 5198-5202.

Lee, P. C. and D. Meisel, 1982: Adsorption and Surface-Enhanced Raman of Dyes on Silver and Gold Sols. *Journal of Physical Chemistry*, **86,** 3391-3395.

Martinez, W. L. and A. R. Martinez, 2001: *Computational Statistics Handbook with MATLAB*. 1 ed. Chapman & Hall/CRC, 616 pp.

McCreery, R. L., 2000: *Raman spectroscopy for chemical analysis*. *Chemical analysis; v. 157.*, Wiley-Interscience, xxiv, 420, [4] of plates pp.

Mullen, K. I., D. X. Wang, L. G. Crane, and K. T. Carron, 1992: Trace detection of ionic species with surface enhanced Raman spectroscopy. *Spectroscopy (Duluth, MN, United States)*, **7,** 24-31.

Pearman, W. F. and A. W. Fountain, III, 2006: Classification of chemical and biological warfare agent simulants by surface-enhanced Raman spectroscopy and multivariate statistical techniques. *Applied Spectroscopy*, **60,** 356-365.

Peters, C. J. and D. M. Hartley, 2002: Anthrax inhalation and lethal human infection. *Lancet*, **359,** 710-1.

Priebe, C. E., 1994: Adaptive Mixtures. *Journal of the American Statistical Society*, **89,** 796-806.

Solka, J. L., W. L. Poston, and E. J. Wegman, 1995: A Visualization Technique for Studying the Iterative Estimation of Mixture Densities. *Journal of Computational and Graphical Statistics*, **4,** 180-198.

Torok, T. J., R. V. Tauxe, R. P. Wise, J. R. Livengood, R. Sokolow, S. Mauvais, K. A. Birkness, M. R. Skeels, J. M. Horan, and L. R. Foster, 1997: A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA: the journal of the American Medical Association*, **278,** 389-95.